

Klasifikasi Data Kelulusan Mahasiswa Menggunakan Algoritma C.45

Azwar Anas^{1*}, Ade Jermawinsyah Zebua², Akhmadi³

Sistem Informasi, Fakultas Pertanian, Sains dan Teknologi, Universitas Graha Karya Muara Bulian¹
Manajemen, Fakultas Hukum, Ekonomi dan Bisnis, Universitas Graha Karya Muara Bulian^{2,3}
Jl. Gajah Mada Muara Bulian - Indonesia
azwarzayn@gmail.com¹, adejermawinsyahzebua9@gmail.com², h.akhmadijambi@gmail.com³

Submitted : 06/08/2025; Reviewed : 11/09/2025; Accepted : 17/10/2025; Published : 31/00/2025

Abstract

The development of information technology and computer science has driven various institutions, including higher education institutions, to optimize the use of data in decision-making processes. Each student possesses a distinct set of data, which varies according to their academic and non-academic abilities. The greater the number of students, the more diverse the range of values that can be observed. These data, however, would merely appear as meaningless collections if no in-depth analysis were conducted. When analyzed using data mining methods, the data can reveal new, previously hidden knowledge. The purpose of this study is to classify student data based on two target outcomes: graduating on time and not graduating on time. The method employed is the C4.5 classification algorithm implemented through the Weka data mining software. The training dataset consists of 83 records. The results show that the model achieved an accuracy rate of 91.5663%, with 61 students predicted to graduate on time out of a total of 83 classified cases. The variable with the highest gain value was the Grade Point Average (GPA), followed by the Scholarship variable.

Keywords: data mining, students, classification, the C4.5 algorithm, outcomes.

Abstrak

Perkembangan teknologi informasi dan ilmu komputer telah mendorong berbagai institusi, termasuk perguruan tinggi, untuk mengoptimalkan pemanfaatan data dalam proses pengambilan keputusan. Setiap mahasiswa memiliki himpunan data yang akan berbeda. Hal ini tergantung dari mahasiswa dalam kemampuan akademik maupun non akademik. Semakin besar jumlah mahasiswa, semakin beragam pula variasi nilai yang dihasilkan. Kumpulan data tersebut akan tampak seperti sekadar tumpukan informasi tanpa makna apabila tidak dilakukan analisis secara mendalam. Namun, melalui penerapan metode data mining, data tersebut dapat diolah menjadi **pengetahuan baru** yang sebelumnya tersembunyi dan tidak teridentifikasi secara eksplisit. Tujuan dari penelitian ini adalah untuk mengklasifikasikan data mahasiswa berdasarkan dua target yaitu lulus tepat waktu dan tidak tepat waktu. Metode yang digunakan adalah algoritma klasifikasi C4.5 dan software data mining Weka. Data training yang digunakan berjumlah 83. Hasil penelitian menunjukkan tingkat akurasi data diklasifikasi dengan benar mencapai 91,5663% dan 61 mahasiswa akan lulus tepat waktu dari total 83 orang yang diklasifikasikan. Variabel dengan gain tertinggi adalah IPK dan diikuti oleh Beasiswa.

Kata kunci: data mining, mahasiswa, klasifikasi, algoritma C4.5, lulusan.

1. Pendahuluan

Perkembangan teknologi informasi dan ilmu komputer telah mendorong berbagai institusi, termasuk perguruan tinggi, untuk mengoptimalkan pemanfaatan data dalam proses pengambilan keputusan. Salah satu bentuk pemanfaatan tersebut adalah melalui penerapan teknik *data mining*, yang bertujuan untuk mengekstraksi pola, informasi tersembunyi, atau pengetahuan baru dari kumpulan data yang besar. Di lingkungan perguruan tinggi, data akademik mahasiswa seperti nilai, indeks prestasi, keaktifan, dan status studi merupakan aset penting yang dapat dianalisis guna mendukung kebijakan akademik yang berbasis data (*data-driven decision making*).

Klasifikasi merupakan salah satu teknik dalam *data mining* yang digunakan untuk memetakan data ke dalam kelas-kelas tertentu berdasarkan atribut-atribut yang dimilikinya [1]. Teknik ini sangat berguna dalam konteks pendidikan, khususnya untuk mengidentifikasi pola karakteristik mahasiswa berdasarkan data historis, sehingga memungkinkan pihak universitas melakukan prediksi terhadap kondisi atau status mahasiswa di masa mendatang. Dengan demikian, klasifikasi dapat digunakan sebagai alat bantu dalam proses evaluasi dan perencanaan strategi pembinaan akademik.

Algoritma C4.5 adalah satu dari beberapa algoritma klasifikasi yang populer dan banyak digunakan karena kemampuannya dalam membangun pohon keputusan (*decision tree*) secara efisien [2]. Algoritma ini mampu menangani atribut data numerik maupun kategorik, serta mengelola data yang tidak lengkap. Dengan menggunakan pendekatan *information gain ratio*, C4.5 secara sistematis membentuk struktur pohon yang dapat digunakan untuk mengklasifikasikan data baru dengan tingkat akurasi yang tinggi.

Penelitian ini bertujuan untuk mengimplementasikan algoritma C4.5 dalam klasifikasi data kelulusan mahasiswa Universitas Graha Karya Muara Bulian (UGKMB). Melalui implementasi ini, diharapkan dapat dihasilkan model klasifikasi yang dapat membantu institusi dalam mengidentifikasi kecenderungan status kelulusan mahasiswa, serta memberikan kontribusi terhadap sistem pendukung keputusan dalam lingkungan pendidikan tinggi.

Setiap mahasiswa memiliki himpunan data yang akan berbeda. Hal ini tergantung dari mahasiswa dalam kemampuan akademik maupun non akademik. Semakin besar jumlah mahasiswa, semakin beragam pula nilai yang dihasilkan. Kumpulan data tersebut akan tampak seperti sekadar tumpukan informasi tanpa makna apabila tidak dilakukan analisis secara mendalam. Namun, ketika dianalisis dengan metode data mining (penambangan data), informasi tersebut dapat diubah menjadi pengetahuan baru yang sebelumnya tersembunyi [3].

Data mining memanfaatkan beragam algoritma yang disesuaikan dengan karakteristik data serta jenis pengetahuan yang hendak dieksplorasi. Istilah *Pattern Recognition* sering digunakan sebagai istilah lain yang merujuk pada bidang kajian ini [4]. Beberapa algoritma yang umum digunakan dalam proses data mining antara lain algoritma K-Means, yang berfungsi untuk melakukan pengelompokan (*clustering*) terhadap data berdasarkan kemiripan karakteristik antar objek [5], algoritma *Market Basket Analysis* (MBA) [6], *Frequent Pattern Tree* [7], dan Apriori [8] untuk menggali kaidah asosiasi (*Association Rule*).

Penelitian ini menerapkan algoritma C4.5 untuk melakukan **klasifikasi** terhadap data mahasiswa UGKMB. Pendekatan ini selaras dengan tujuan penelitian, yaitu memperoleh pengetahuan yang bermakna dari kumpulan data yang sebelumnya belum terstruktur atau tersembunyi.

Formula yang digunakan dalam algoritma C4.5 [9]:

$$\text{Entropy (S)} = \sum_{i=1}^n (-p_i) * \log_2(p_i)$$

S = jumlah kasus

n = jumlah partisi atribut S

p_i = proporsi dari s_i terhadap S

$$\text{Gain (S,A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{entropy (s}_i)$$

$|S_i|$ = jumlah kasus pada partisi ke-i

$|S|$ = jumlah kasus dalam S

Beberapa penelitian terkait diantaranya:

1. Penelitian pada analisis pengadaan peralatan kantor menggunakan algoritma C4.5 berhasil membantu pengambil Keputusan dalam menentukan pengadaan alat kantor yang efektif [10].
2. Penelitian tentang klasifikasi gizi balita dengan Tingkat akurasi mencapai 95,50% [11].
3. Penelitian tentang perbandingan algoritma Naïve Bayes dan C4.5 dengan hasil lebih tinggi Tingkat akurasi Nive Bayes dibandingkan C4.5 [12].
4. Penelitian tentang klasifikasi penyakit stroke dengan hasil Tingkat akurasi mencapai 99,07% [13].
5. Penelitian tentang klasifikasi kualitas udara DKI Jakarta yang focus pada faktor yang mempengaruhi kualitas udara [14].

2. Metode Penelitian

Penelitian ini adalah penelitian pustaka (*library research*). Penulis menggunakan algoritma C4.5 karena merupakan komponen dari algoritma data mining yang sesuai dengan jenis data dan pengetahuan yang akan diproses. Jumlah data training yang digunakan adalah 83 mahasiswa dan data testing sebanyak 20. Weka adalah software data mining yang digunakan.

Agar menghasilkan penelitian yang berkualitas, setiap peneliti perlu menerapkan metode yang tepat dan relevan dengan tujuan penelitian. Proses penelitian diharapkan berlangsung secara sistematis dan ilmiah sesuai dengan tahapan metode yang telah dirancang. Adapun kerangka penelitian yang digunakan dapat dijelaskan sebagai berikut.



Gambar 1. Kerangka Penelitian

Penelitian yang dilakukan sebagaimana Langkah-langkah berikut :

1. Menetapkan Ruang Lingkup Penelitian
Untuk menjamin bahwa penelitian tetap terfokus pada permasalahan yang dikaji, langkah awal yang perlu dilakukan adalah menetapkan **ruang lingkup** serta **batasan masalah** agar arah penelitian menjadi jelas dan terarah.
2. Analisa Masalah Penelitian
Analisis masalah dilakukan untuk mengidentifikasi dan menetapkan permasalahan penelitian beserta batasannya. Pada tahap ini, algoritma **C4.5** diterapkan untuk melakukan **klasifikasi data mahasiswa**, sehingga pola dan hubungan antar variabel dapat diidentifikasi secara lebih terstruktur.
3. Penentuan Tujuan
Setelah masalah dipahami, tujuan penelitian ini ditetapkan. Tujuannya adalah untuk menjawab masalah-masalah sebelumnya. Seperti menggali tumpukan data mahasiswa untuk kemudian dihasilkan suatu pengetahuan.
4. Kajian Literatur
Untuk mencapai tujuan penelitian, berbagai referensi yang terkait dengan topik penelitian diperiksa. Referensi ini kemudian dipilih sesuai dengan kebutuhan penelitian dan dipilih dari jurnal ilmiah, buku, dan sumber internet yang relevan.
5. Pengumpulan Data
Untuk memahami masalah yang ada, proses pengumpulan data dimulai dengan melakukan observasi langsung pada objek penelitian. Studi kepustakaan dilakukan dengan membaca berbagai buku yang relevan dengan masalah penelitian. Data utama yang digunakan adalah mahasiswa.
6. Penetapan Teknik yang digunakan
Penetapan **teknik pengolahan data** merupakan tahapan penting dalam proses analisis data. Pada penelitian ini, penulis menerapkan **algoritma C4.5** sebagai metode utama untuk melakukan pengolahan dan analisis terhadap data yang telah dikumpulkan.

7. Implementasi Sistem
Selanjutnya, implementasi dilakukan pada *software data mining* Weka.
8. Pengujian Sistem
Untuk membandingkan hasil **perhitungan manual** dengan hasil **komputerisasi menggunakan perangkat lunak**, dilakukan tahap **pengujian sistem** dengan prosedur sebagai berikut:
 - a. Melakukan pengujian data mahasiswa menggunakan formula algoritma C4.5 secara manual untuk memperoleh hasil awal klasifikasi.
 - b. Menerapkan algoritma C4.5 sebagai metode analisis dengan memanfaatkan perangkat lunak Weka untuk mengolah data secara komputerisasi.
 - c. Membandingkan hasil pengolahan manual dengan hasil pengolahan melalui *software*. Apabila perbedaan hasil yang diperoleh tidak signifikan, maka hasil analisis dapat dianggap valid dan akurat.

3. Hasil Penelitian

3.1. Data Mahasiswa

Penelitian ini menggunakan data mahasiswa yang telah pada Program Studi Manajemen UGKMB. Adapun variabelnya yaitu Nama Mahasiswa, IPK, SKS, Kehadiran, Beasiswa, Organisasi dan Status Kelulusan. Secara khusus nama mahasiswa pada tabel hanya akan ditampilkan huruf abjad sebagai bentuk penyamaran identitas guna menjaga kerahasiaan data pribadi responden dan tidak diproses dalam algoritma C4.5.

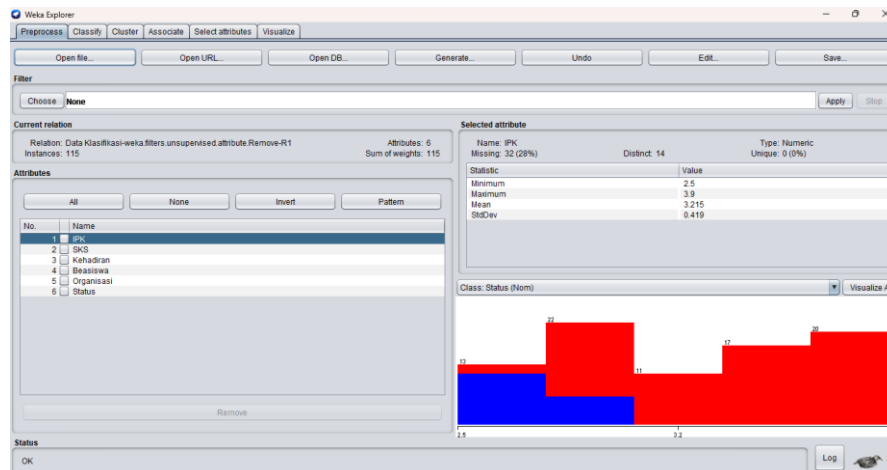
Tabel 1. Data Mahasiswa

No.	Nama	IPK	SKS	Kehadiran	Beasiswa	Organisasi	Status Kelulusan
1	A	02.05	126	Baik	Tidak	Ya	Tidak Tepat Waktu
2	B	3	137	Kurang	Tidak	Ya	Tepat Waktu
3	C	03.04	134	Baik	Tidak	Ya	Tepat Waktu
4	D	0,136111	126	Cukup	Tidak	Ya	Tepat Waktu
5	E	0,145139	128	Cukup	Ya	Ya	Tepat Waktu
6	F	03.04	132	Cukup	Ya	Tidak	Tepat Waktu
7	G	3	123	Baik	Tidak	Tidak	Tepat Waktu
8	H	03.08	127	Baik	Tidak	Tidak	Tepat Waktu
9	I	03.09	124	Kurang	Tidak	Tidak	Tepat Waktu
10	J	03.01	120	Kurang	Tidak	Ya	Tepat Waktu
11	K	02.54	139	Baik	Ya	Tidak	Tepat Waktu
12	L	0,178472	132	Cukup	Tidak	Tidak	Tepat Waktu
13	M	03.04	123	Cukup	Tidak	Tidak	Tepat Waktu
14	N	03.21	139	Cukup	Tidak	Tidak	Tepat Waktu
15	O	3	123	Baik	Tidak	Tidak	Tepat Waktu
16	P	0,171528	126	Baik	Tidak	Tidak	Tepat Waktu
17	Q	03.36	140	Baik	Tidak	Tidak	Tepat Waktu
18	R	0,141667	128	Baik	Tidak	Ya	Tidak Tepat Waktu
19	S	02.05	121	Kurang	Ya	Ya	Tidak Tepat Waktu
20	T	3	140	Baik	Ya	Ya	Tepat Waktu
21	U	03.04	140	Baik	Ya	Ya	Tepat Waktu
22	V	0,136111	121	Baik	Tidak	Ya	Tidak Tepat Waktu
23	W	0,145139	136	Baik	Tidak	Ya	Tidak Tepat Waktu
24	X	03.04	121	Cukup	Tidak	Tidak	Tepat Waktu
25	Y	3	128	Cukup	Ya	Tidak	Tepat Waktu
26	Z	03.08	123	Baik	Tidak	Tidak	Tepat Waktu
27	AA	03.09	130	Baik	Tidak	Tidak	Tepat Waktu
83	AB	03.01	134	Baik	Tidak	Tidak	Tepat Waktu

Tabel 1 di atas menyajikan data mahasiswa dengan atribut IPK, SKS, Kehadiran, Beasiswa, Organisasi, dan Status Kelulusan sebagai target. Data tersebut akan diproses lebih lanjut dengan menggunakan fungsi algoritma C4.5.

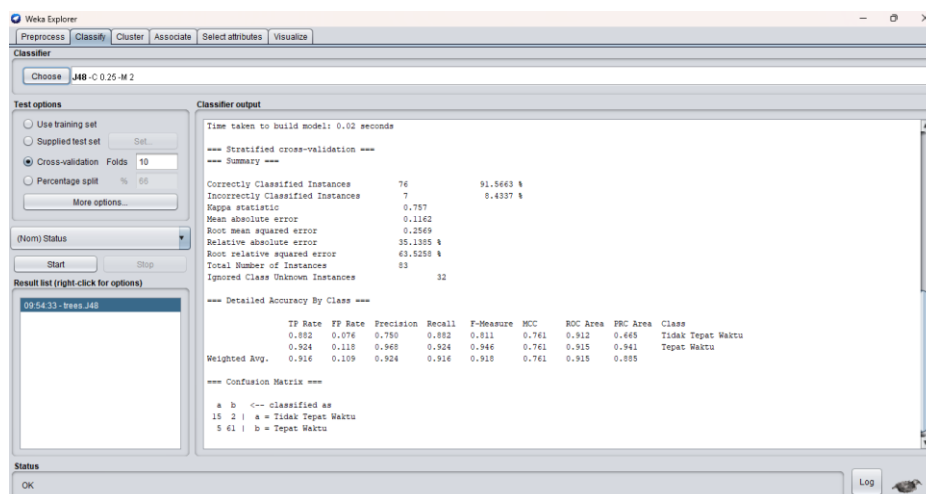
3.2. Analisis Algoritma C4.5

Data pada tabel 1 di atas, terlebih dahulu dibuat dalam format CSV (comma delimited). Tampilkan data pada software Weka sebagaimana gambar berikut.



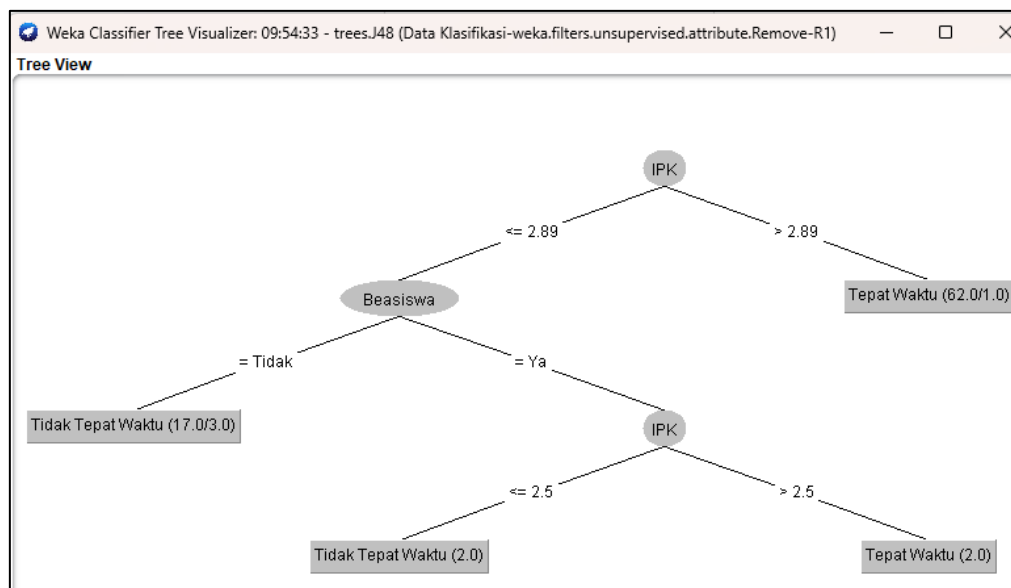
Gambar 2. Input Data Mahasiswa

Pada gambar 2 di atas, enam variabel ditampilkan dan satu variabel target yaitu status, yang akan menjadi Keputusan apakah lulus tepat waktu atau tidak tepat waktu. Kemudian pilih menu Classify untuk menerapkan fungsi klasifikasi. Terdapat beberapa algoritma, namun dalam hal ini pilih J48 sebagai implementasi dari algoritma C4.5, sebagaimana gambar berikut.



Gambar 3. Implementasi Klasifikasi Algoritma C4.5

Pada gambar di atas, pada bagian correctly classified instance sebesar 91,5663% yang menunjukkan tingkat akurasi data klasifikasi. Kappa statistic sebesar 0,757 merupakan alat ukur kesepakatan prediksi dan kebenaran, semakin mendekati 1 (satu) maka semakin baik. Pada bagian confusion matrix terlihat bahwa 15 (lima belas) orang mahasiswa tidak tepat waktu dan masuk klasifikasi tidak tepat waktu (benar). 2 (dua) orang tidak tepat waktu namun masuk klasifikasi tepat waktu (salah). 5 (lima) orang tidak tepat waktu namun masuk klasifikasi tepat waktu (salah). 61 (enam puluh satu) orang akan tepat waktu dan masuk klasifikasi tepat waktu (benar). Kemudian buat pohon Keputusan berdasarkan gambar sebelumnya sebagaimana gambar berikut.



Gambar 4. Pohon Keputusan Algoritma C4.5

Berdasarkan gambar 4 diatas, terlihat variabel IPK sebagai cabang pertama diikuti dengan variabel beasiswa. Jika IPK $\geq 2,89$ maka akan tepat waktu menyelesaikan studinya. Jika IPK $\leq 2,89$ dan tidak mendapat beasiswa, maka tidak akan tepat waktu menyelesaikan studinya. Jika mendapat beasiswa dan IPK $\leq 2,5$ maka tidak akan tepat waktu menyelesaikan studinya. Jika mendapat beasiswa dan IPK $> 2,5$ maka akan tepat waktu menyelesaikan studinya.

Terdapat dua variabel yang paling berpengaruh terhadap kelulusan tepat waktu mahasiswa yaitu IPK dan beasiswa. IPK memiliki pengaruh besar karna hal ini menggambarkan kemampuan akademik mahasiswa selama mengikuti proses perkuliahan, hal ini juga sejalan dengan syarat kelulusan mahasiswa baik di tingkat mata kuliah maupun institusi. Sedangkan beasiswa memberikan mahasiswa fokus dalam proses perkuliahan tanpa perlu mengalihkan pikirannya tentang biaya perkuliahan, sehingga secara tidak langsung berdampak pada prestasi akademik mahasiswa sebagaimana terlihat pada IPK tersebut.

4. Penutup

Berdasarkan pembahasan di atas, dengan menggunakan data training berjumlah 83 orang berhasil mengklasifikasikan mahasiswa baik tepat waktu maupun yang tidak tepat waktu menyelesaikan studinya. Tingkat akurasi yang tinggi yakni mencapai 91,5663% terbukti dari 61 (enam puluh satu) orang mahasiswa diprediksi akan tepat waktu menyelesaikan studinya dan masuk kategori benar. Artinya mahasiswa tersebut hampir dipastikan menyelesaikan studinya tepat waktu. Sementara 15 (lima belas) orang dipastikan tidak akan tepat waktu menyelesaikan studinya.

Penelitian selanjutnya perlu dilakukan perbandingan beberapa algoritma sejenis seperti Naïve Bayes atau Random Forest untuk melihat Tingkat akurasi yang lebih tinggi dan dengan penggunaan data yang lebih besar, sehingga dapat diimplementasikan secara riil pada perguruan tinggi.

Daftar Pustaka

- [1] Yusuf Maulana, Riki Winanjaya, and Fitri Rizki, "Penerapan Data Mining dengan Algoritma C4.5 Dalam Memprediksi Penjualan Tempe," *Bull. Comput. Sci. Res.*, vol. 2, no. 2, pp. 53–58, 2022, doi: 10.47065/bulletincsr.v2i2.163.
- [2] I. Anggraeni and S. Andriani, "3562-9606-1-Pb," vol. 18, no. 2, pp. 62–68, 2021.
- [3] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma," *JUKOMIKA (Jurnal Ilmu Komput. dan Inform.)*, vol. 4, no. 2, pp. 57–71, 2022, doi: 10.54650/jukomika.v4i2.368.
- [4] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk

- Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [5] N. T. Suswanto, P. Chyan, and V. Putri, “Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City,” vol. 11, no. 4, 2021.
- [6] A. Anas, Azwar; Jermawinsyah Zebua, “Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat,” *J. Ilm. MEDIA SISFO*, vol. 16, no. 1, pp. 54–61, 2022, doi: 10.35957/jatisi.v7i2.195.
- [7] B. S. Hasugian, “Penerapan Metode Association Rule Untuk Menganalisa Pola Pemakaian Bahan Kimia Di Laboratorium Menggunakan Algoritma FP-Growth (Studi Kasus di Laboratorium Kimia PT . PLN (Persero) Sektor Pembangkitan Belawan Medan) Buyung Solihin Hasugian Universitas,” *J. Ilmu Komput. dan Inform.*, vol. 6341, no. November, pp. 56–69, 2019.
- [8] D. Ai, H. Pan, X. Li, Y. Gao, and D. He, “Association rule mining algorithms on high-dimensional datasets,” *Artif. Life Robot.*, vol. 23, no. 3, pp. 420–427, 2018, doi: 10.1007/s10015-018-0437-y.
- [9] A. Prasetyo and A. Wahyono, “Implementasi Algoritma C4.5 dalam Klasifikasi Kondisi Ekonomi Warga Kabupaten Boyolali,” *Gener. J.*, vol. 8, no. 2, pp. 121–131, 2024, doi: 10.29407/gj.v8i2.22827.
- [10] A. Iftitah and R. Setyadi, “Penerapan Algoritma C4.5 Untuk Analisis Pengadaan Peralatan dan Mesin Kantor,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 434–442, 2023, doi: 10.47065/josh.v4i2.2673.
- [11] R. A. A. Sri Lestari, “Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatuk Kabupaten Garut Jawa Barat,” *J. Wahana Pendidik.*, vol. 8, no. 10, pp. 116–125, 2022.
- [12] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, and Shelly Janu Setyaning Tyas, “Analisis Perbandingan Algoritma Naive Bayes Dan C4.5 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan,” *Tematik*, vol. 8, no. 1, pp. 96–103, 2021, doi: 10.38204/tematik.v8i1.576.
- [13] R. E. Pambudi, Sriyanto, and Firmansyah, “Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5,” *Ijccs*, vol. x, No.x, no. x, pp. 1–5, 2022.
- [14] M. Astriyani, I. N. Laela, D. P. Lestari, L. Anggraeni, and T. Astuti, “Analisis Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma C4.5,” *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 6, no. 1, pp. 36–41, 2023, doi: 10.32524/jusitik.v6i1.790.